

# Spacing between core recognition motifs determines relative orientation of AraR monomers on bipartite operators

Deepti Jain and Deepak T. Nair\*

Laboratory 4, National Centre for Biological Sciences (NCBS-TIFR), UAS-GKVK Campus, Bellary Road, Bangalore, Karnataka 560065, India

Received June 25, 2012; Revised and Accepted September 24, 2012

## ABSTRACT

Transcription factors modulate expression primarily through specific recognition of cognate sequences resident in the promoter region of target genes. AraR (*Bacillus subtilis*) is a repressor of genes involved in L-arabinose metabolism. It binds to eight different operators present in five different promoters with distinct affinities through a DNA binding domain at the N-terminus. The structures of AraR-NTD in complex with two distinct operators (ORA1 and ORR3) reveal that two monomers bind to one recognition motif (T/ANG) each in the bipartite operators. The structures show that the two recognition motifs are spaced apart by six bases in cases of ORA1 and eight bases in case of ORR3. This increase in the spacing in the operators by two base pairs results in a drastic change in the position and orientation of the second monomer on DNA in the case of ORR3 when compared with ORA1. Because AraR binds to the two operators with distinct affinities to achieve different levels of repression, this observation suggests that the variation in the spacing between core recognition motifs could be a strategy used by this transcription modulator to differentially influence gene expression.

## INTRODUCTION

The ability to perceive stimuli and respond appropriately is the hallmark of any living organism. In many instances, such responses are regulated at the level of transcription. A valuable and informative system to study transcription regulation involves modulators that are sensitive to the presence of small metabolites. AraR is one such transcription factor, which regulates carbon catabolism in *Bacillus subtilis*. In the absence of L-arabinose, AraR represses

transcription of approximately 13 genes required for arabinose utilization. DNAase I footprinting and mutation experiments have shown that AraR binds to eight different operator sites within five different promoters present in the *araABDLMNPQ-abfA* operon plus *araR*, *araE*, *abnA* and *xsa* genes (1,2,3). It has been shown that binding of AraR to operator pairs ORA1–ORA2 and ORE1–ORE2 is cooperative and promotes DNA looping (1–3). The loop formation prevents transcription initiation and exerts tight repression. In contrast, binding to ORR3 shows no co-operativity and does not result in DNA distortion (3). Consequently, this mode of repression is less effective and allows basal transcription of *araR* gene.

In the presence of L-arabinose, repression of the genes regulated by AraR is abolished. The transcription of metabolic operon (*araABDLMNPQ-abfA*) and *araE* gene increases by 50-fold and that of *araR* gene by 4-fold (3–6). The consensus sequence recognized by AraR is palindromic 5'ATTTGTACGTACAAAT3' and is 16 bp in length (1). Although there is high sequence similarity between the eight operators recognized by AraR, the repressor binds to each operator with distinct affinities ranging from 40 nM (ORR3) to 250 nM (ORA1) (1).

AraR comprises of two domains, the smaller N-terminus domain (NTD; 1–68 residues) and a larger C-terminus domain (CTD; 71–362). The CTD is the receptor for L-arabinose and is termed as the effector domain (8). The structure of this domain in complex with L-arabinose was determined recently and shows that the CTD mediates homodimerization of AraR (9). Random and site-directed mutagenesis experiments and *in vivo* effects of amino acid substitution have shown that the interaction of AraR with DNA is mediated by the NTD (7,8). We have determined crystal structures of AraR-NTD (ArRNTD) in complex with two different operators, ORA1 (5'-AAAATTGTTTCGTACAAAT ATT-3') and ORR3 (5'-AAATTTGTCCGTATACAT TTT-3') (Table 1 and Supplementary Figure S1). The structures show that each monomer binds to one core recognition motif present twice in the bipartite operators and

\*To whom correspondence should be addressed. Tel: +91 80 2366 6405; Fax: +91 80 2363 6662; Email: deepaknair@ncbs.res.in

**Table 1.** Data collection and Refinement statistics

	ArRNTD <sub>ORR3</sub>		ArRNTD <sub>ORA1</sub> , native	ArRNTD <sub>mORR3</sub> , native
	SeMet	Native		
Wavelength	0.97864 Å	1.0 Å	1.0 Å	1.0 Å
Space group	P4 <sub>3</sub>	P4 <sub>3</sub>	C2	C2
Cell constants (Å)	46, 46, 171.9	46.2, 46.2, 171.2	138.8, 42.4, 67.4	137.8, 42.6, 67.4
(°)	90, 90, 90	90, 90, 90	90, 114.7, 90	90, 114.9, 90
Resolution (Å) <sub>b</sub>	2.79 (2.9–2.79) Å <sup>a</sup>	2.3 (2.47–2.3)	2.3 (2.42–2.3)	1.97 (2.08–1.97)
R <sub>sym</sub> or R <sub>merge</sub> <sup>b</sup>	7.2 (38)	5.0 (27.9)	9.0 (37.8)	7.3 (4.1)
I/σI	19.7(6.0)	14.1 (3.6)	6.9 (2.4)	9.9 (3.0)
Completeness (%)	99.5(96.5)	99.0 (100)	97.6 (97.0)	97.4 (96.1)
Redundancy	10.7 (10.2)	3.9 (3.8)	3.0 (3.0)	4.6 (4.6)
Refinement				
Resolution (Å)		40–2.30	36.0–2.30	40–1.97
No. reflections		15 670	15 886	24 593
R <sub>work</sub> <sup>d</sup> /R <sub>free</sub> <sup>c</sup>		21.9/27.1	22.2/26.2	21.9/25.3
No. atoms				
Protein		1281	1181	1181
DNA		855	855	855
Water		145	103	156
Average B-factors (Å <sup>2</sup> )				
Protein: A		62.2	51.5	33.5
Protein: B		63.8	46.2	38.4
DNA		65.6	48.9	36.2
Water		65.5	46.2	38
Rms deviations				
Bond lengths (Å)		0.010	0.01	0.013
Bond angles (°)		1.208	1.27	1.35

<sup>a</sup>Highest resolution shell is shown in parentheses.

<sup>b</sup>R<sub>merge</sub> = Σ|I - <I>|/ΣI, where I is the integrated intensity of a given reflection.

<sup>c</sup>R<sub>free</sub> was calculated using 5% of data excluded from refinement.

<sup>d</sup>R<sub>work</sub> = Σ||F<sub>obs</sub>| - |F<sub>calc</sub>||/Σ|F<sub>obs</sub>|.

also provide a basis for the observed difference in affinities for the two operators. Comparison of the two complexes shows that the spacing between the two core recognition motifs is a critical determinant of the relative orientation of the two monomers on DNA. The implications of this observation for repression by AraR and transcription regulation in general are discussed.

## MATERIALS AND METHODS

### ArRNTD cloning, expression and purification

The gene segment corresponding to ArRNTD (residues 1–68) was amplified from genomic DNA (Bangalore Genie) and cloned using EcoR1 and Not1 sites, into a modified pET21b plasmid (Novagen). The gene segment was fused to an upstream six-Histidines-tag with a PreScission-cleavable linker (pDJN1). The fusion protein was expressed in C41DE3 cells induced with 0.8 mM isopropyl-D-thiogalactopyranoside (IPTG) at 18°C. Harvested cells were resuspended in lysis buffer (25 mM Tris pH 8.0, 500 mM NaCl and 5% glycerol) and lysed by sonication. Cell lysate was clarified by centrifugation and was loaded onto a HisTrap column (GE Healthcare). Protein was eluted using a step gradient to 100% elution buffer containing 1 M Imidazole. ArRNTD was further purified by gel filtration (Superdex 75, GE Healthcare). The purified protein was concentrated by centrifugal filtration (Argos Technologies) to 54 mg/

ml and flash frozen and stored at –80°C. To prepare selenomethionine-labeled ArRNTD, the protein was expressed in B834 strain of *Escherichia coli* (methionine auxotroph), and the cells were grown using a Se-Met media kit (Molecular Dimensions). Purification and crystallization of the selenomethionyl-ArRNTD was identical to native protein, except buffers included 5 mM dithiothreitol.

### Nucleic acid preparation

Four oligonucleotides corresponding to complementary sequences for the ORA1 and ORR3 (with a T/A overhang) were purchased from Keck Biotechnology Resource Laboratory (Yale University). These oligonucleotides were purified by ion-exchange chromatography (monoQ column) and then desalted (column) and lyophilized. Next, the oligonucleotides were dissolved in appropriate volumes of autoclaved and filtered (0.22 μm) water to achieve a final concentration of 4 mM. Equimolar amounts of complementary oligonucleotides were annealed by heating to 90°C for 5 min followed by cooling to 25°C, and the final concentration of the duplex DNA was 2 mM. Two oligonucleotides corresponding to the modified ORR3 (mORR3) operator were purchased from Sigma Genosys in a purified form, dissolved in sterile autoclaved water and then annealed to obtain a final concentration of 2 mM for duplex DNA.

## Crystallization

Complex co-crystals were obtained using vapor diffusion by mixing the duplex DNA and ArRNTD (1–68) in the ratio 1.2:1, with the final concentration of DNA at 0.67 mM. The mixture was incubated on ice for 20 min. The complex crystallized in solutions of PEG 8 K (20%) buffered with 0.1 M sodium acetate (pH = 4.5) and containing 200 mM KCl. Crystals were cryoprotected by serial transfers using increasing concentrations of glycerol from 5% to 25% in steps of 5%. The crystals were flash frozen in liquid nitrogen. Crystallization condition for selenomethionine-labeled ArRNTD in complex with ORR3 was identical to that for the complex with native protein. Also, crystallization condition and the cryoprotection strategy for the ArRNTD<sub>ORA1</sub> and ArRNTD<sub>mORR3</sub> complexes were identical to that for ArRNTD<sub>ORR3</sub>.

## Structure determination

The ArRNTD<sub>ORR3</sub> complex structure was solved by single-wavelength anomalous diffraction (SAD) method using crystals prepared from SeMet-labeled protein. SAD data were collected at peak wavelength (0.97864) of the X-ray absorption spectrum (Table 1) at the European Synchrotron Radiation facility (beamline BM14), Grenoble. In addition, data from native crystals of ArRNTD<sub>ORR3</sub>, ArRNTD<sub>ORA1</sub> and ArRNTD<sub>mORR3</sub> (PXIII, Swiss Light Source) were collected to a maximal resolution of 2.3, 2.3 and 1.97 Å, respectively (Table 1). The data were processed using HKL2000 (10). For the ArRNTD<sub>ORR3</sub> complex, using the anomalous signal from SeMet, one of the possible two Se sites in the asymmetric unit was located using ShelxD (11). An initial experimental electron density map could be computed using ShelxE using anomalous data up to 3.0 Å. Using this map, the polypeptide chain corresponding to ArRNTD could be built in Coot (12). Next, the DNA model was built using the make-na server (<http://structure.usc.edu/make-na/server.html>) and was docked into the electron density. The map was improved through iterative cycles of rigid body refinement in CNS (13). The model obtained was used for molecular replacement in MOLREP (14) against the native data set and gave a unique and unambiguous solution. This was followed by iterative cycles of refinement [CNS/PHENIX (15)] and model building (Coot) with constant monitoring of geometrical parameters. The final round of refinement was performed in REFMAC incorporating TLS restraints (16,17). For the final model, the  $R_{\text{free}}$  is 27.1% and  $R_{\text{work}}$  is 21.9% (Table 1).

The ArRNTD<sub>ORA1</sub> and ArRNTD<sub>mORR3</sub> complex structures were solved by molecular replacement in PHASER using one molecule of ArRNTD and double-stranded DNA as the model from the ArRNTD<sub>ORR3</sub> and ArRNTD<sub>ORA1</sub> structures, respectively. The second molecule of ArRNTD was docked into the electron density, and map was improved through iterative cycles of rigid body refinement in CNS. Using Coot, the DNA sequence of ORR3 was modified to match that of ORA1 and mORR3 sequences. This was followed by cycles of

model building (Coot) refinement (CNS and PHENIX) and water picking (PHENIX). Final rounds of refinement were performed in REFMAC incorporating TLS restraints. The  $R_{\text{free}}$  and  $R_{\text{work}}$  converged to final values of 26.2/22.2% and 25.3/21.9% for the ArRNTD<sub>ORA1</sub> and ArRNTD<sub>mORR3</sub>, respectively (Table 1).

Electron density maps of the protein–DNA interface for ArRNTD<sub>ORA1</sub> and ArRNTD<sub>ORR3</sub> complexes are shown in Supplementary Figure S1. For the final refined models of the three complexes, PROCHECK (18) revealed that 98% of the residues of the NTD are in the allowed regions of the Ramachandran plot. The structure was analyzed using the CONTACT (CCP4 suite), and area of the surface buried at the interface was calculated using CNS. The bend in the operator DNAs was calculated using CURVES (19) program, and shape complementarity index ( $Sc$ ) was calculated using the SC program (CCP4 suite) (17,20).

## Data deposition

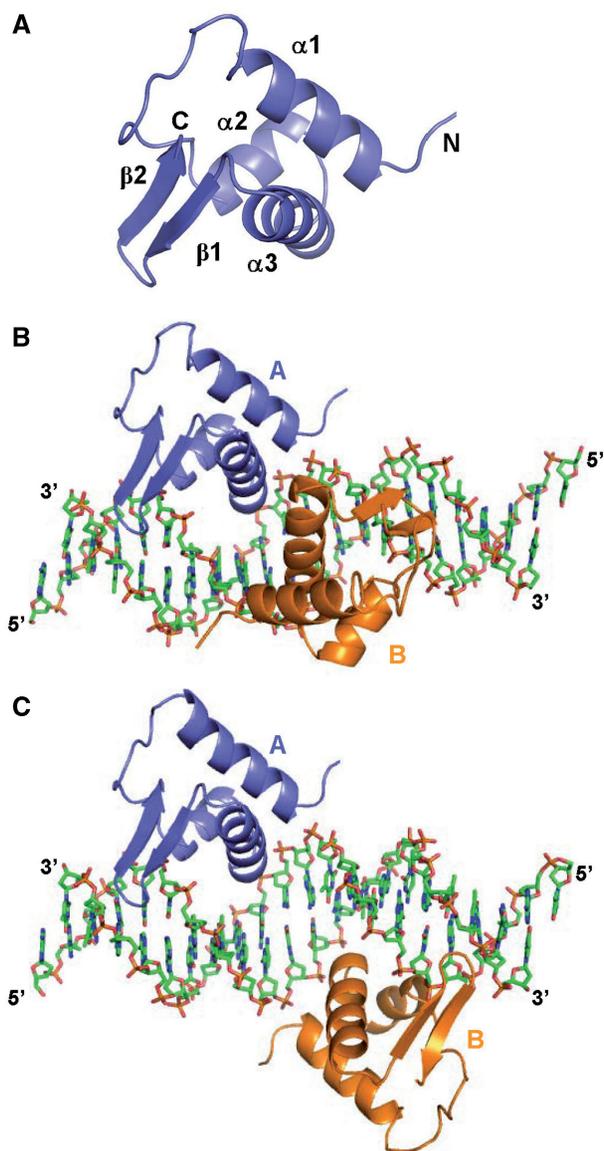
Atomic coordinates and structure factors have been deposited in the PDB with accession codes 4EGY (ArRNTD<sub>ORA1</sub>), 4EGZ (ArRNTD<sub>ORR3</sub>) and 4H0E (ArRNTD<sub>mORR3</sub>).

## RESULTS

### Overall structure of ArRNTD<sub>ORA1</sub> and ArRNTD<sub>ORR3</sub>

We have determined crystal structures of ArRNTD in complex with two different natural operators, ORA1 (5'-AAAATTGTTTCGTACAAATATT-3') and ORR3 (5'-AAATTTGTCCGTATACATTTT-3') (Table 1 and Supplementary Figure S1). The DNA duplexes in the ArRNTD<sub>ORA1</sub> and ArRNTD<sub>ORR3</sub> complexes contain a T/A overhang and pack head to tail to form a pseudo continuous double helix. The duplexes used span from –5 to +15 with respect to the transcription start sites of the respective promoters. In both the complexes, there are two monomers of ArRNTD (monomers A and B) and one molecule of double-stranded DNA in the asymmetric unit (Figure 1). There is no interaction between the two monomers of ArRNTD in either of the complexes (Figure 1B and C). The structures of the two monomers bound to DNA are nearly identical. Monomer A aligns with monomer B with rmsds of 0.25 Å and 0.19 Å for the ArRNTD<sub>ORA1</sub> and ArRNTD<sub>ORR3</sub> complexes (68 C $\alpha$  atoms), respectively. In addition, the structure of ArRNTD is identical in both ArRNTD<sub>ORA1</sub> and ArRNTD<sub>ORR3</sub> complexes. The superimposition of monomer A and monomer B of ORA1 onto the corresponding monomers of ORR3 yields rmsds of 0.37 Å and 0.43 Å (68 C $\alpha$  atoms), respectively.

ArRNTD monomers show the presence of the characteristic winged helix-turn-helix (HTH) motif (Figure 1A). Each monomer shows three consecutive alpha helices  $\alpha$ 1 (4–19),  $\alpha$ 2 (30–38) and  $\alpha$ 3 (41–54) followed by a  $\beta$ -sheet comprising of two strands  $\beta$ 1 (56–61) and  $\beta$ 2 (63–67). The HTH motif is formed by  $\alpha$ 2 and  $\alpha$ 3 and the latter forms the recognition helix and is oriented towards the major groove. The  $\beta$ -sheet represents the wing domain and



**Figure 1.** Structure of ArRNTD-operator DNA complex. (A) Structure of a monomer of ArRNTD shown in ribbon representation with different secondary structural elements labeled. (B) Structure of ArRNTD<sub>ORA1</sub> complex. Protein is shown as ribbon representation and monomer A colored slate blue and monomer B colored orange. DNA is shown in stick representation and the 5'- and 3'-ends are labeled. (C) Structure of ArRNTD<sub>ORR3</sub>. The figure is color coded as in (B).

residues in the loop (residues 60–63) in the  $\beta$ -sheet form a type II  $\beta$ -turn, which is inserted in the adjacent minor groove (Figure 1B and C).

#### ArRNTD: ORA1 interactions

The structure of the ArRNTD<sub>ORA1</sub> complex shows that the interactions of monomer A with DNA are similar to interactions of monomer B. Hence, a description of key interactions is restricted to monomer A (Figure 2A and C). Base-specific interactions are formed between residues of monomer A with nucleotides Ade3' (Q61:A) and Gua5

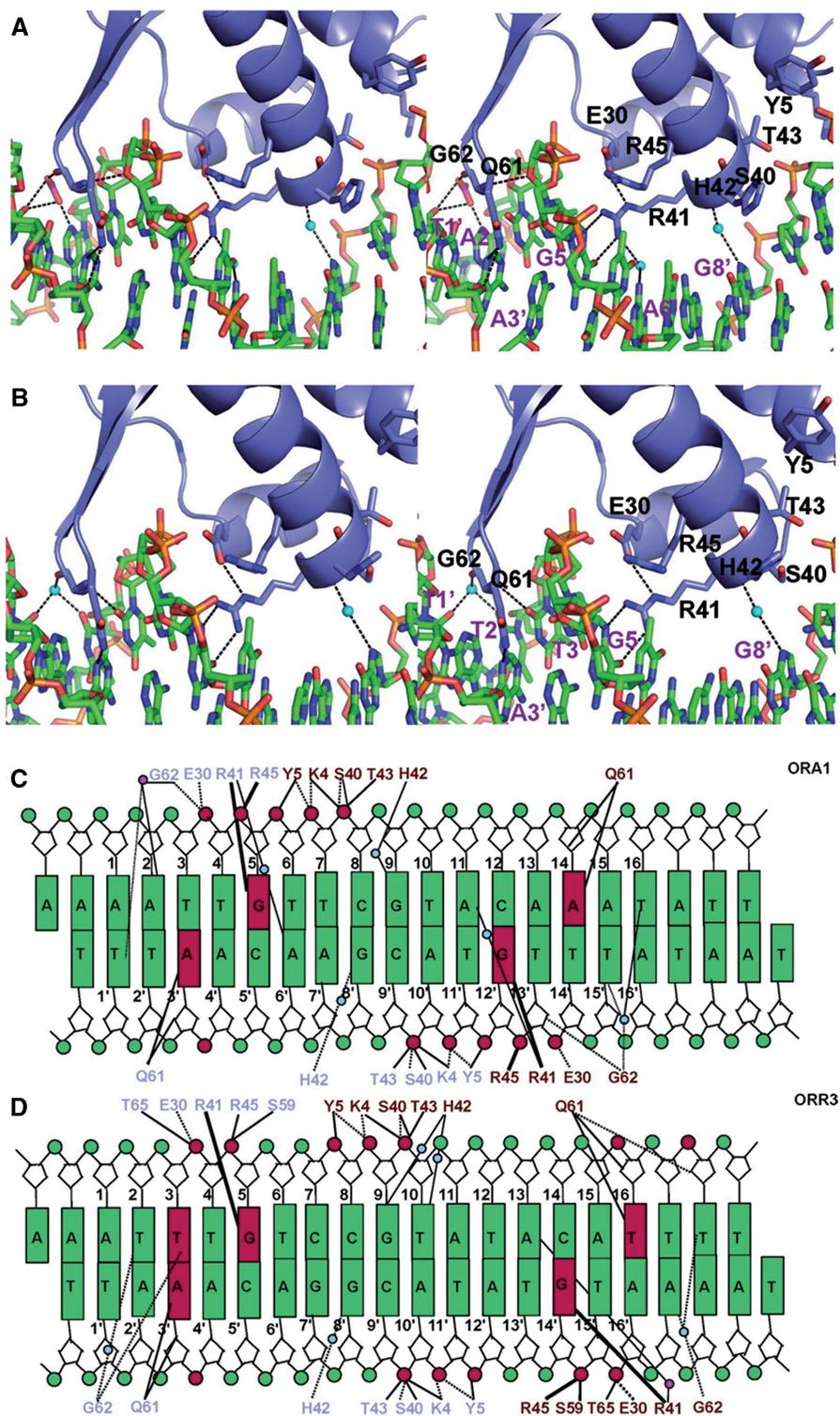
(R41:A). R41:A is part of the recognition helix  $\alpha$ 3 and forms a bidentate hydrogen bond with the Hoogsteen edge of Gua5 (Figure 2A). The importance of this observed interaction is supported by the fact that the mutation Gua5 $\rightarrow$ Thy in the ORA1 has an adverse effect on the affinity for AraR *in vitro* and also reduces repression by AraR *in vivo* (7). Q61:A is present in the  $\beta$ -turn of the wing motif that is inserted into the minor groove and interacts with Ade3'. This interaction is essential for function as the mutation Q61 $\rightarrow$ A completely abolished repression activity *in vivo* and drastically decreased DNA-binding ability *in vitro* (7). Identical interactions are seen for corresponding residues of monomer B with G12' (R41:B) and A14 (Q61:B) (Figure 2C).

The key interaction that defines specificity of AraR towards cognate DNA is that of R41 with the guanine base. Surprisingly, the R41 $\rightarrow$ A mutation does not have a drastic effect on repression *in vivo*. However, the mutation of the R45 (present in close proximity to R41)  $\rightarrow$  L/A results in complete loss of regulatory activity (7,8). It is seen that both R41 and R45 interact with residue E30 (Figure 2A and B). The side chain of E30 forms a hydrogen bond with R41 and holds the Arg side chain in position to interact with DNA. In addition, E30 also forms a hydrogen bond with R45, which in turn contacts the DNA phosphate backbone at Gua5. E30 appears to be instrumental in orienting both R45 and R41 to interact with DNA. Consistent with the observed critical role of this residue, the mutation E30 $\rightarrow$ A reduces repression *in vivo* 5-fold and affects affinity towards ORA1 *in vitro* by an order of magnitude (7). As a result of their close proximity, it is possible that the guanidinium group of R45 can swing in and substitute for that of R41 in the mutant R41 $\rightarrow$ A and thus rescue function.

Additional base-specific interactions in the form of water-mediated hydrogen bonds are present for Ade6' (R41:A), Gua8' (H42:A), Gua9 (H42:B), Ade11 (R41:B), Thy15' (G62:B) and Thy16 (G62:B) (Figure 2A and C). These water-mediated interactions are important for productive operator binding as the mutations Gua8' $\rightarrow$ Thy, Gua9 $\rightarrow$ Thy/Ade/Cyt, Ade6' $\rightarrow$ Cyt, Ade11 $\rightarrow$ Thy/Gua/Cyt and Thy16 $\rightarrow$ Gua in ORA1 adversely affect DNA binding (tested *in vitro*) and transcriptional repression (tested *in vivo*) (7). Consistent with the observed interactions for H42, H42 $\rightarrow$ A substitution in AraR led to a decrease in repression *in vivo* and affinity toward ORA1 *in vitro* (7). In addition to water-mediated interactions, an acetate ion forms bridging interactions between G62:A and the bases Ade2 and Thy1' (Figure 2A and C). Also, the two monomers of ArRNTD show similar interactions and direct sequence recognition through base-specific interactions are largely restricted to two TNG recognition motifs (AAAATTGTTCGTACAAATATT) present within the ORA1 operator (Figure 2).

#### ArRNTD: ORR3 interactions

As in the case of ArRNTD<sub>ORA1</sub>, monomer A of ArRNTD<sub>ORR3</sub> forms base-specific interactions with Ade3' (Q61:A) and Gua5 (R41:A) (Figure 2B and D).



**Figure 2.** Protein–DNA interactions. Stereoview of the direct and indirect base-specific interactions formed between AraR and DNA in (A) ArRNTD<sub>ORA1</sub> and (B) ArRNTD<sub>ORR3</sub> complexes. The interactions of monomer A are shown in the two complexes. Side chains of the interacting protein residues are displayed. Direct and water-mediated hydrogen bonds (<3.5Å) are shown with dashed lines, and the water molecules are shown as blue spheres. The acetate ion is displayed in stick representation. Schematic representation of the protein–DNA interactions is shown for the (C) ArRNTD<sub>ORA1</sub> and (D) ArRNTD<sub>ORR3</sub> complexes. The interacting protein residues from monomer A and monomer B are displayed in slate blue and dark red, respectively. The continuous lines represent the hydrogen bonds between protein side chains and DNA. The dash lines denote the hydrogen bonds with main chain of the protein, and the bold lines represent more than one hydrogen bond with the same residue. Blue and magenta circles represent the water molecules and acetate ions, respectively.

However, in the case of monomer B, the equivalent interactions are with Thy16 (Q61:B) and Gua14' (R41:B) (Figure 2D). In ArRNTD<sub>ORA1</sub>, the interacting nucleotides for monomer B were Ade14 and Gua12'. Thus, there is an increase in the spacing between the two recognition motifs (TNG and ANG) within ORR3 by two base pairs (Figure 2C and D). In the case of monomer B of ORR3, the interaction of R41:B with the nucleotide corresponding to Gua12', i.e. Gua14' is conserved. In ORR3, at a position analogous to that of Ade3', Thy16 is present in the other recognition site (Ade14 in ORA1) but Q61:B still forms a base-specific interaction with the thymine base (Figure 2D). Thus, Q61 exhibits plasticity as far as the identity of the interacting base (purine or pyrimidine) is concerned. An additional base-specific interaction is seen in ArRNTD<sub>ORR3</sub> with Thy3 (G62:A) (Figure 2B and D). An acetate ion forms bridging interactions between R41:B with Ade13 (Figure 2D). Also, base-specific recognition through water-mediated interactions are observed in case of Thy2(G62:A), Gua8'(H42:A), Gua9(H42:B), Thy10(H42:B), Ade17'(G62:B) and Thy18(G62:B) (Figure 2B and D). G62 is located in the minor groove and forms both direct and water-mediated base-specific interactions (Figure 2). This position cannot be occupied by any other residue than glycine as the presence of a side chain will lead to steric clashes with the DNA atoms.

In both ArRNTD<sub>ORR3</sub> and ArRNTD<sub>ORA1</sub> complexes, stabilizing hydrogen bonds are formed with the backbone atoms of DNA by residues K4, Y5 and T43 (Figure 2). The non-specific interactions formed by K4 and Y5 are important for function as the mutations K4→A and Y5→F affect repression by AraR up to 30-fold and 3.3-fold, respectively (7). Overall, the comparison of the two complexes shows that critical base-specific interactions formed by R41 and Q61 are present in all four engagements of ArRNTD with DNA (Figure 2C and D). Consequently, direct sequence recognition by AraR is largely restricted to the core T/ANG recognition motif present within the two half sites of each operator (Figures 1 and 3).

#### Comparison of the structures of ArRNTD<sub>ORA1</sub> and ArRNTD<sub>ORR3</sub> complexes

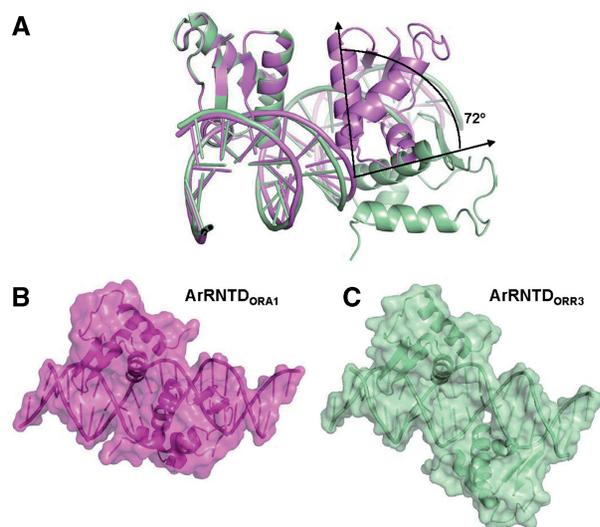
In both the complexes, the area of the interface between each ArRNTD monomer and DNA is approximately the same (~1750 Å<sup>2</sup>). The protein and DNA surfaces in the two complexes exhibit similar shape complementarity (Sc values of 0.68) with that for the B monomer of ArRNTD<sub>ORR3</sub> being marginally higher (0.69). However, the spacing between the core recognition motifs is six bases in the case of ORA1 and eight bases in the case of ORR3. As a result, the relative orientation of the two monomers on DNA is significantly different in the two complexes (Figures 1B and C and 3A). Alignment of the two structures by superimposition of the A monomers shows that monomer B in ArRNTD<sub>ORR3</sub> spins along the DNA axis further away from monomer A, such that the angle between the  $\alpha$ 3 helix of the B monomers is 72° (Figure 3A). This corresponds well with the observed rotation of 36° per base for B-DNA.

An additional outcome of the increase in spacing is the difference in the extent to which the helical axis of DNA is bent on ArRNTD binding. The bend is 6.5° in case of ArRNTD<sub>ORA1</sub> but increases to 20° in the ArRNTD<sub>ORR3</sub> complex. Although the interactions with the core recognition motif are largely conserved, ArRNTD monomers form more direct hydrogen bonds with ORR3 than with ORA1. There are 34 hydrogen bonds formed between protein and DNA residues in ArRNTD<sub>ORR3</sub> when compared with 30 in ArRNTD<sub>ORA1</sub>. The direct interactions unique to the ArRNTD<sub>ORR3</sub> complex are largely through residues present in the wing of monomer B. It appears that the increased spacing in ArRNTD<sub>ORR3</sub> improves the grip of the two monomers on DNA resulting in the increased bend of the helical axis and greater number of polar interactions leading to higher affinity. Another important effect of the increase in spacing is the presentation of distinct surfaces in the two nucleoprotein complexes (Figure 3B and C). The Connolly surfaces of the two complexes shows non-uniform features due to the difference in the orientation of monomer B and the differences in the bend of the DNA helical axis.

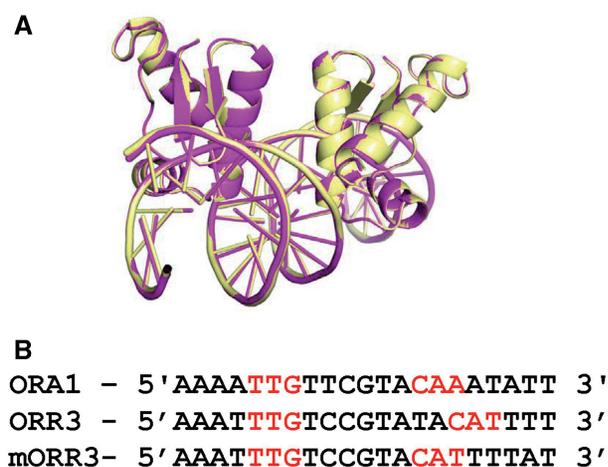
To verify whether the new position of the second monomer is a consequence of the increase in the spacing between the two core recognition motifs, we designed a variant of the ORR3 operator called mORR3. In this modified operator, two nucleotides Ade and Thy at positions 11 and 12 were removed between the core recognition motifs to give rise to a spacing that is equal to that in the case of ORA1 (six bases). The complex ArRNTD<sub>mORR3</sub> crystallized in the space group C2 and X-ray diffraction data could be collected to a maximal resolution of 1.97 Å. The structure of the ArRNTD<sub>mORR3</sub> complex (Table 1) was identical to that of the ArRNTD<sub>ORA1</sub> complex and the two structures superimpose with a rmsd of 0.181 Å for 136 C $\alpha$  atoms (Figure 4). The position of monomer B is identical to that seen in case of the ArRNTD<sub>ORA1</sub> complex. As expected, the interactions between protein and DNA in the two structures are almost identical with the only difference arising due to the presence of Thy at position 14 in mORR3 when compared with Ade in the case of ORA1. Consequently, in the ArRNTD<sub>mORR3</sub> complex, Q61 interacts with the O<sub>2</sub> atom of thymine. Overall, this structure proves that the difference in relative orientation of monomer B between ArRNTD<sub>ORA1</sub> and ArRNTD<sub>ORR3</sub> is a direct consequence of the different spacing between the core recognition motifs in these two operators.

#### Comparison with FadR<sub>FadB</sub> complex

The NTD of AraR (ArRNTD) belongs to the GntR family of transcription regulators (21). The GntR family is further subdivided into seven subfamilies (FadR, HutC, MocR, YtrA, DevA, PlmA and AraR), depending on the topologies of the effector binding C-terminal domain (22,23). The only member of the GntR family for which the crystal structure is available in complex with DNA is FadR (fatty acid degradation regulator) (24,25). The DNA-binding domain of FadR recognizes cognate operator FadB (CATCTGGTACGACCAGATC)



**Figure 3.** Relative orientation of monomers on DNA. (A) Structural alignment of the ArRNTD–DNA complexes. Monomers A from ArRNTD<sub>ORA1</sub> and ArRNTD<sub>ORR3</sub> complexes have been superimposed. The ArRNTD<sub>ORA1</sub> complex is shown in magenta, and the ArRNTD<sub>ORR3</sub> complex is shown in green color. The angle between the recognition helix of the B-monomer in the two complexes is shown. The surface of the ArRNTD monomers when bound to DNA is displayed for (B) ArRNTD<sub>ORA1</sub> and (C) ArRNTD<sub>ORR3</sub>.



**Figure 4.** Comparison of ArRNTD<sub>ORA1</sub> and ArRNTD<sub>mORR3</sub>. (A) Structural alignment of ArRNTD<sub>ORA1</sub> (magenta) and ArRNTD<sub>mORR3</sub> (yellow) complexes. (B) Alignment of the DNA sequences used for crystallization of ArRNTD<sub>ORA1</sub> (ORA1), ArRNTD<sub>ORR3</sub> (ORR3) and ArRNTD<sub>mORR3</sub> (mORR3). The recognition motif in each case is colored red.

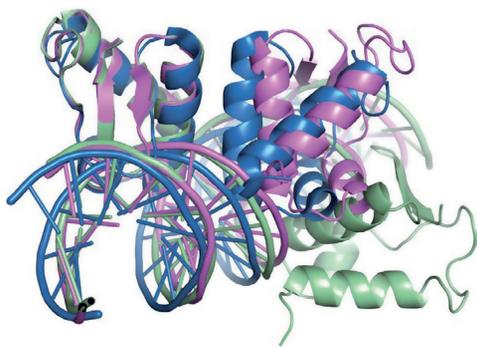
through a winged helix-turn-helix motif in a manner similar to AraR. The rmsd for superimposition of the NTD of monomer A of FadR (residues 5–72) on monomer A in the ArRNTD<sub>ORR3</sub> complex is 1.9 Å, (68 C $\alpha$  atoms). R45 in FadR forms a bidentate hydrogen bond with the Hoogsteen edge of Gua5 (similar to R41 in case of ArRNTD) (24,25). Thus, the recognition of Gua5 through a bidentate hydrogen bond with Arg present in  $\alpha$ 3 seems to be a conserved feature of this family. In the FadR<sub>FadB</sub> complex, direct base-specific protein–DNA

interactions occur mainly through the TGG recognition motif present in each half site (CATCTGGTACGA CCAGATC). The NTD of members of the GntR family is predicted to bind a signature sequence 5'-(N)<sub>y</sub>GT(N)<sub>x</sub>AC(N)<sub>y</sub>-3' where the number *x* and *y* vary (22,23). On the basis of the base-specific interactions observed in structures of the ArRNTD<sub>ORA1</sub>, ArRNTD<sub>ORR3</sub> and FadR<sub>FadB</sub> complexes, the signature sequence can be modified to 5'-(N)<sub>y</sub>TNG-(N)<sub>x</sub>CNA/T-(N)<sub>y</sub>-3'.

Although there is similarity in the overall structure for each monomer of ArRNTD and FadR-NTD bound to DNA, the position of monomer B on DNA with respect to monomer A shows divergence in the two cases (Figure 5). The spacing between the two recognition motifs is 5 bp in the case of FadR<sub>FadB</sub> complex. Consequently, the two monomers are closer to each other than in case of ArRNTD<sub>ORA1</sub> and ArRNTD<sub>ORR3</sub> and  $\alpha$ 3 of the two monomers interact with each other to form a dimer interface. Overall, a comparison of the three complexes shows that the position and orientation of monomer B with respect to monomer A changes as the spacing between the core recognition motifs varies.

## DISCUSSION

The ArRNTD<sub>ORA1</sub> and ArRNTD<sub>ORR3</sub> structures show that the relative orientation of two monomers of a dimeric transcription factor on DNA is a function of the extent of spacing between corresponding recognition motifs in a bipartite operator. The importance of spacing between core binding sites on structure and function has been seen in case of two transcriptional regulators Pit-1 and Oct-1 (26,27). These regulators show the presence of two domains POUh and POUl connected by a flexible linker that are both able to bind to distinct sites on DNA. Pit-1 plays a critical role in ensuring cell-type-specific restriction of growth hormone expression (26). This is primarily achieved through distinct modes of interaction with two bipartite sites that show differences in the spacing between the two half sites. The POUh and POUl domains of Pit-1 are organized differently on binding to the GH-1 site (spacing: 4 bp) and Prl-P1 (spacing: 6 bp). This allows the protein to recruit repressor complex in case of GH-1 site and possibly an activator complex in case Prl-P1 (26). Similarly, for a related transcription factor Oct-1, the POUh and POUl domains exhibit distinct modes of organization on two different cognate sequences—termed MORE and PORE—that allow presentation of distinct protein surfaces probably for recruitment of unique downstream effectors (27). Transcription factors such as p53 and RFX-1 that are known to regulate transcription by binding to a number of different operators wherein the two half sites are separated by spacers of different lengths (28,29). In case of the activator p53, variable spacing in different cognate sequences has been shown to influence affinity of binding to cognate sites (30) and could result in different levels of activation for different genes.



**Figure 5.** Comparison with the FadR<sub>FadB</sub> complex. The structures were aligned through superimposition of monomer A from the ArRNTD<sub>ORA1</sub> (magenta), ArRNTD<sub>ORR3</sub> (green) and FadR<sub>FadB</sub> (blue) complexes.

Based on the observations made in case of Pit-1, Oct-1 and p53, it is possible that the two different binding modes of ArRNTD to ORA1 and ORR3 might be exploited to generate different levels of repression. Although there are no interactions between the NTD monomers bound to DNA, the full length protein is expected to bind to the symmetric cognate operators as a dimer formed through interactions between the effector binding domain (8,9). Mota and coworkers (1,3) have shown that binding of AraR to the ORA1–ORA2 and the ORE1–ORE2 pairs is cooperative and ultimately results in the formation of a DNA loop to achieve tight repression. The affinity of AraR for individual operators ORA1 (250 nM) and ORA2 (>250 nM) increases by approximately 8- and >5-fold when the two operators are present together (K<sub>d</sub> of 34 nM and 47 nM for ORA1 and ORA2, respectively). Also, in the case of ORE1 and ORE2, the affinity for the individual operators increases by 3- and 2 fold with a decrease in K<sub>d</sub> from 108 and 127 nM to 35 and 55 nM, respectively, when the two operators are present together. The authors suggested that co-operative binding and looping occur because the spacing between the ORA1 and ORA2 operators (42 bp) and that between ORE1 and ORE2 operators (43 bp) will place the bound AraR dimers in phase. It is expected that this would allow the formation of stabilizing contacts between the effector domains of AraR dimers bound at each operator. This interaction would lead to an increase in the affinity of binding at the two sites and also facilitate looping. In case of ORR3, the nearest AraR operator ORE1 is 81 bp upstream. Mota and coworkers (1,3) suggest that this places the dimers bound to ORE1 and ORR3 out of phase thereby preventing looping. However, studies probing repression by LacR show that an inter-operator spacing of 81 bp does not present a barrier to looping *in vivo* (31,32).

The structures presented here raise the possibility that when the two AraR monomers are closer to each other (as in case of ORA1 complex), they attain a dimer organization through their effector domains that presents a distinct molecular surface to facilitate contacts between AraR dimers bound to ORA1–ORA2 or ORE1–ORE2 and hence promote DNA looping. On the other hand, when the two AraR monomers are bound further apart on the

operator (as in case of ORR3), the dimer organization of the effector domains is such that interaction with other dimers bound at proximal operators ORE1 is not possible. As a result, looping and the consequent enhanced repression will not happen, and this allows basal transcription of *araR* gene. In summary, the increased intra-operator spacing observed in case of ORR3 may help avoid adventitious stringent repression of the *araR* gene. It is vital to allow baseline expression of the *araR* gene as strong repression would disrupt the regulatory network that prevents expression of corresponding metabolic genes in the absence of arabinose.

A number of transcription factors are known to bind to DNA in a similar manner as AraR- one monomer/dimer binds each one of the two recognition motifs in a symmetric operator. The structures of ArRNTD in complex with ORA1 and ORR3 suggest that condensation in the form of oligomers on double helical DNA allows these transcription factors to amplify small changes in the linear DNA sequence into large changes in three dimensional structure. The difference in the structures probably leads to presentation of distinct surfaces for further oligomerization through space or the recruitment of different transcription factors to differentially influence and regulate gene expression. The structures also show that the presentation of the linear sequence of DNA in a double helical form adds another layer to the information resident in DNA. This additional layer can be read and exploited by the same transcription factor to achieve functionally diverse outcomes.

#### ACCESSION NUMBERS

4EGY, 4EGZ and 4H0E.

#### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Figure 1.

#### ACKNOWLEDGEMENTS

D.T.N. thanks the X-ray diffraction facility located in the Molecular Biophysics Unit of the Indian Institute of Science [funded by Department of Biotechnology (DBT) and Department of Science and Technology (DST), Government of India] for facilitating screening and data collection. D.T.N. acknowledges the help rendered by Dr. Hassan Belrhali (ESRF) and Dr. Babu Manjashetty during data collection at the BM-14 beamline of ESRF and by Dr. Meitian Wang during data collection at the PXIII beamline of SLS. Figures were generated using PyMOL (Schroedinger Inc.). The authors thank Prof. Jayant Udgaonkar (NCBS) for critically reading the manuscript.

#### FUNDING

Department of Atomic Energy, Government of India; Department of Science & Technology [Sanction Order

number: SR/FT/LS-122/200, for 3 years starting from 24 September 2009]; Ramanujan fellowship from the Department of Science and Technology, Government of India (to D.T.N.); Young Investigator Grant from Department of Science and Technology, Government of India (to D.J.); the BM14 project—a collaboration between Department of Biotechnology, Government of India, EMBL and ESRF (for data collection at the BM14 beamline of ESRF [Grenoble, France]). Funding for open access charge: Department of Atomic Energy, Government of India.

*Conflict of interest statement.* None declared.

## REFERENCES

- Mota, L.J., Tavares, P. and Sa-Nogueira, I. (1999) Mode of action of AraR, the key regulator of L-arabinose metabolism in *Bacillus subtilis*. *Mol. Microbiol.*, **33**, 476–489.
- Raposo, M.P., Inacio, J.M., Mota, L.J. and de Sa-Nogueira, I. (2004) Transcriptional regulation of genes encoding arabinan-degrading enzymes in *Bacillus subtilis*. *J. Bacteriol.*, **186**, 1287–1296.
- Mota, L.J., Sarmiento, L.M. and de Sa-Nogueira, I. (2001) Control of the arabinose regulon in *Bacillus subtilis* by AraR in vivo: crucial roles of operators, cooperativity, and DNA looping. *J. Bacteriol.*, **183**, 4190–4201.
- Sa-Nogueira, I. and Mota, L.J. (1997) Negative regulation of L-arabinose metabolism in *Bacillus subtilis*: characterization of the araR (araC) gene. *J. Bacteriol.*, **179**, 1598–1608.
- Sa-Nogueira, I., Nogueira, T.V., Soares, S. and de Lencastre, H. (1997) The *Bacillus subtilis* L-arabinose (ara) operon: nucleotide sequence, genetic organization and expression. *Microbiology*, **143**(Pt 3), 957–969.
- Sa-Nogueira, I. and Ramos, S.S. (1997) Cloning, functional analysis, and transcriptional regulation of the *Bacillus subtilis* araE gene involved in L-arabinose utilization. *J. Bacteriol.*, **179**, 7705–7711.
- Franco, I.S., Mota, L.J., Soares, C.M. and de Sa-Nogueira, I. (2007) Probing key DNA contacts in AraR-mediated transcriptional repression of the *Bacillus subtilis* arabinose regulon. *Nucleic Acids Res.*, **35**, 4755–4766.
- Franco, I.S., Mota, L.J., Soares, C.M. and de Sa-Nogueira, I. (2006) Functional domains of the *Bacillus subtilis* transcription factor AraR and identification of amino acids important for nucleoprotein complex assembly and effector binding. *J. Bacteriol.*, **188**, 3024–3036.
- Prochazkova, K., Cermakova, K., Pachel, P., Sieglöva, I., Fabry, M., Otwinowski, Z. and Rezacova, P. (2012) Structure of the effector-binding domain of the arabinose repressor AraR from *Bacillus subtilis*. *Acta Crystallogr. D Biol. Crystallogr.*, **68**, 176–185.
- Z. Otwinowski and Minor, M., (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology*, **276**: C.W. Carter, Jr. & R. M. Sweet, Eds., Academic Press (New York) 307–326.
- Sheldrick, G.M. (2008) A short history of SHELX. *Acta Crystallogr. A*, **64**, 112–122.
- Emsley, P. and Cowtan, K. (2004) Coot: model building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
- Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S. *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.
- Vagin, A. and Teplyakov, A. (1997) MOLREP and automated program for molecular replacement. *J. Appl. Crystallogr.*, **30**, 1022–1025.
- Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
- Winn, M.D., Murshudov, G.N. and Papiz, M.Z. (2003) Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol.*, **374**, 300–321.
- Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 235–242.
- Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, **26**
- Lavery, R., Moakher, M., Maddocks, J.H., Petkeviciute, D. and Zakrzewska, K. (2009) Conformational analysis of nucleic acids revisited: curves+. *Nucleic Acids Res.*, **37**, 5917–5929.
- Lawrence, M.C. and Colman, P.M. (1993) Shape complementarity at protein/protein interfaces. *J. Mol. Biol.*, **234**, 946–950.
- Haydon, D.J. and Guest, J.R. (1991) A new family of bacterial regulatory proteins. *FEMS Microbiol. Lett.*, **63**, 291–295.
- Rigali, S., Derouaux, A., Giannotta, F. and Dusart, J. (2002) Subdivision of the helix-turn-helix GntR family of bacterial regulators in the FadR, HutC, MocR, and YtrA subfamilies. *J. Biol. Chem.*, **277**, 12507–12515.
- Hoskisson, P.A. and Rigali, S. (2009) Chapter 1: Variation in form and function the helix-turn-helix regulators of the GntR superfamily. *Adv. Appl. Microbiol.*, **69**, 1–22.
- Xu, Y., Heath, R.J., Li, Z., Rock, C.O. and White, S.W. (2001) The FadR-DNA complex. Transcriptional control of fatty acid metabolism in *Escherichia coli*. *J. Biol. Chem.*, **276**, 17373–17379.
- van Aalten, D.M., DiRusso, C.C. and Knudsen, J. (2001) The structural basis of acyl coenzyme A-dependent regulation of the transcription factor FadR. *EMBO J.*, **20**, 2041–2050.
- Scully, K.M., Jacobson, E.M., Jepsen, K., Lunyak, V., Viadiu, H., Carriere, C., Rose, D.W., Hooshmand, F., Aggarwal, A.K. and Rosenfeld, M.G. (2000) Allosteric effects of Pit-1 DNA sites on long-term repression in cell type specification. *Science*, **290**, 1127–1131.
- Remenyi, A., Tomilin, A., Pohl, E., Lins, K., Philippsen, A., Reinbold, R., Scholer, H.R. and Wilmanns, M. (2001) Differential dimer activities of the transcription factor Oct-1 by DNA-induced interface swapping. *Mol. Cell*, **8**, 569–580.
- Horvath, M.M., Wang, X., Resnick, M.A. and Bell, D.A. (2007) Divergent evolution of human p53 binding sites: cell cycle versus apoptosis. *PLoS Genet.*, **3**, e127.
- Gajiwala, K.S., Chen, H., Cornille, F., Roques, B.P., Reith, W., Mach, B. and Burley, S.K. (2000) Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature*, **403**, 916–921.
- Kitayner, M., Rozenberg, H., Kessler, N., Rabinovich, D., Shaulov, L., Haran, T.E. and Shaked, Z. (2006) Structural basis of DNA recognition by p53 tetramers. *Mol. Cell*, **22**, 741–753.
- Muller, J., Barker, A., Oehler, S. and Muller-Hill, B. (1998) Dimeric lac repressors exhibit phase-dependent co-operativity. *J. Mol. Biol.*, **284**, 851–857.
- Zhang, Y., McEwen, A.E., Crothers, D.M. and Levene, S.D. (2006) Analysis of in-vivo LacR-mediated gene repression based on the mechanics of DNA looping. *PLoS One*, **1**, e136.